

GFP Gene Targeting System

The components of the system we used to study gene targeting are depicted in Fig. S1. It is based on the correction by homologous recombination of a mutated green fluorescent protein (GFP) gene (“GFP gene targeting system”). A single copy of an artificial gene target (A658) is stably integrated into the genome of 293 cells, creating cell line “293/A658.” The genomic target consists of a GFP gene that is mutated by the insertion of an in-frame stop codon followed by the recognition site for the *I-SceI* endonuclease (“Sce”) at basepair (bp) 327 of the GFP coding region. 293/A658 cells remain GFP negative (data not shown). We measured gene targeting by transfecting 293/A658 with a “substrate” construct that was promoter-less and missing the first 37 nucleotides, including the initiation codon, of the GFP gene (“RS2100” for repair substrate with 2100 bp of homology). After transfection we measured the number of GFP positive cells by flow cytometry and determined the rate of gene targeting by normalizing the percentage of GFP positive cells to the transfection efficiency. When we transfected RS2100 into 293 cells without a gene target (293-0 cells) there were no GFP positive cells (data not shown). When we transfected supercoiled RS2100 alone into 293/A658, the gene targeting rate was 7.1×10^{-7} , or 0.71 events per million transfected cells (“spontaneous gene targeting or “S-GT”) (Fig. S1, B and C). We found that the frequency of S-GT increased as the insertional mutation size in the GFP gene decreased (Fig. S1C).

We measured the double-stranded break induced gene targeting (DSB-GT) rate by co-transfecting RS2100 with a Sce expression plasmid (see Fig. S1B for a representative flow cytometry plot). Creating a DSB in the target GFP gene increased the rate of gene targeting over 2000-fold (Fig. S1C). Transfection of the Sce expression plasmid without

the repair substrate RS2100 generated no GFP positive cells (Fig. S1B for a representative flow cytometry plot). Transfection of the Sce expression plasmid with the repair substrate RS2100 into 293-0 cells did not generate GFP positive cells (data not shown). Unlike spontaneous gene targeting, the DSB-GT did not decrease as the insertional mutation size increased (Fig. S1C). We determined the time course of DSB-GT by serially measuring the number of GFP positive cells after transfection and found that the maximum number of GFP positive cells occurred 2.5-3 days after transfection and remained stable for at least two weeks (Fig. S1D). The gene targeting reaction, therefore, occurred relatively quickly after transfection and created stable genetic change. Finally, we purified single GFP positive cells by fluorescence activated cell sorting and determined the sequence of the GFP gene after gene targeting. In the 9 GFP positive cells tested, the sequence of the GFP gene was wild-type, demonstrating that gene targeting had occurred (data not shown). The DSB-GT rate in murine 3T3 and human SaOS-2 cells was similar to that in 293 cells (data not shown). Further, the DSB-GT rate was similar whether a pool of cells or a clonal cell line with single insertion sites for A658 was examined (data not shown). Thus, our findings were not cell type or integration site dependent.

Using the GFP gene targeting system we quantitatively explored several variables that regulate the rate of DSB-GT. We found that increasing the amount of substrate (RS2100) transfected increased the rate of DSB-GT until a plateau is reached (Fig. S2A). This result demonstrated that gene targeting is dependent on the amount of repair substrate available. We found that increasing the length of homology between the repair substrate and the target linearly increased the rate of DSB-GT (Fig. S2B). In these

experiments we kept the amount of 5' homology constant at 290 basepairs (bp) and varied the amount of 3' homology from 500 bp to 3700 bp. This result suggests that while spontaneous gene targeting is logarithmically dependent on homology length (*SI*), DSB-GT is linearly dependent on homology length. In either case, increasing the length of homology between the target and the repair substrate increased the frequency with which the cell undergoes gene targeting. Figure S2C shows that the DSB-GT rate was linearly dependent on the amount of PGK-Sce transfected. The DSB-GT rate reached a plateau, however, when higher amounts of CBA-Sce were transfected (Fig. S2C). This data suggests that DSB-GT is dependent on the creation of a DSB to initiate gene targeting but eventually becomes saturated for DSB creation. We found that manipulating the transcriptional status of the repair substrate can affect the rate of gene targeting. Transcribing the truncated repair substrate with a CMV promoter (CMV-RS2100) increased the rate of DSB-GT by 50% (Fig. S2D). Just as with RS2100, transfecting CMV-RS2100 into 293-0 cells did not generate GFP positive cells (data not shown). The rate of DSB-GT was highest when Sce expression is driven by the CBA promoter, intermediate with the CMV promoter, and lowest with the PGK promoter (Fig. S2E). This result probably reflects the different levels of Sce expression from each promoter. Figure 1E also demonstrates that the rate of DSB-GT can be increased by placing the repair substrate on the same plasmid as the Sce expression cassette rather than co-transfecting two plasmids. The stimulation was lost when the DSB-GT process was at saturation, as when the CBA promoter was used to express Sce. When we optimized the above parameters we achieved gene targeting rates of 3-5% which is a 20-fold increase

over the DSB-GT rate obtained without optimization and gives an overall stimulation of gene targeting by a DSB of over 40,000-fold.

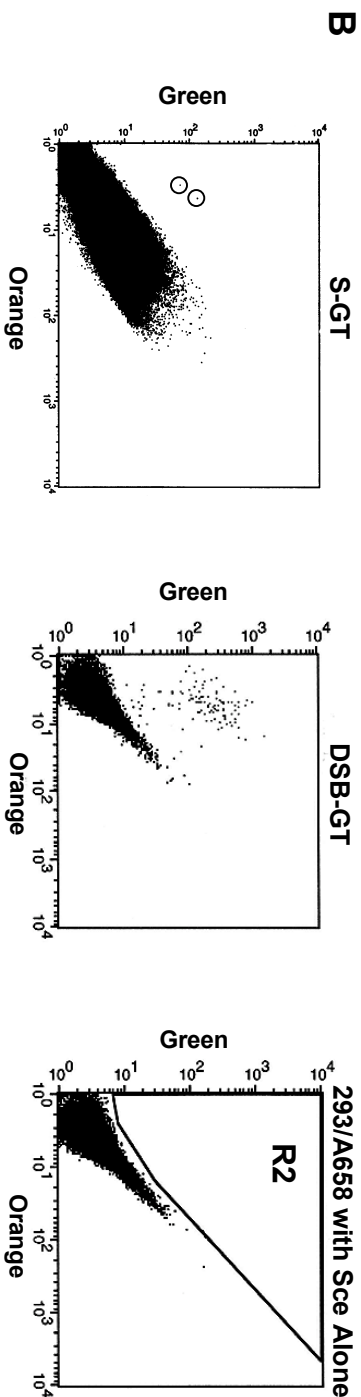
Supplementary Figure 1

A Gene Target (A658): CMV/CBA GFP* IRES CD8 WPRE PGK-NEO

Stop-Sce site

Repair Substrate (RS2100): 37 truncGFP IRES CD8

Repair Substrate (RS2700): 37 truncGFP IRES CD8 WPRE



C

Gene Target	1bp insertion	7 bp insertion	35 bp insertion (A658)	35 bp insertion (A658)	66 bp insertion (QQR8)	Fold Stimulation By Sce Induced DSB
Sce	No	No	No	Yes	Yes	
Events/ 10 ⁶ cells	29.5 +/- 13.0	1.5 +/- 1.1	0.71 +/- 0.27	1600 +/- 500	1900 +/- 250	>2000
Rate	2.9 x 10 ⁻⁵	1.5 x 10 ⁻⁶	7.1 x 10 ⁻⁷	1.6 x 10 ⁻³	1.9 x 10 ⁻³	

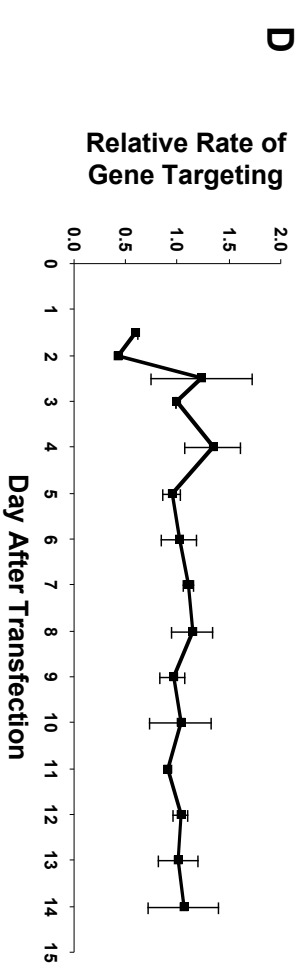


Fig. S1. GFP gene targeting system. A) Gene Targeting System. The artificial gene target (A658) consisted of a GFP gene mutated by a 35 basepair insertion which includes a stop codon and a recognition site for the *I-SceI* endonuclease (Sce) (5' TAGGGATAACAGGGTAAT 3') at basepair 327 of the coding sequence. The GFP gene was driven by a hybrid cytomegalovirus enhancer/chicken β -actin promoter ("CMV/CBA" or "CBA"). The GFP gene was part of bicistronic transcript in which an internal ribosomal entry site ("IRES") allowed translation of the human CD8 α gene ("CD8"). The bicistronic message contained a Woodchuck post-transcriptional regulatory element ("WPRE") to increase messenger RNA levels (S2). Finally, the locus contained a gene with the phosphoglycerate kinase promoter ("PGK") driving the neomycin phosphotransferase gene (NEO) to allow selection by the antibiotic G418. The repair substrates RS2100 and RS2700 are also depicted. They consisted of a GFP gene that has been truncated at basepair 37 of the coding sequence and thus were missing the initiation codon ("truncGFP"). The truncated GFP gene was followed by the IRES-CD8 for RS2100 or IRES-CD8-WPRE for RS2700 as in A658. The A658 target gene was introduced into 293 cells by electroporating 2×10^6 cells with 10 μ g of supercoiled A658 plasmid DNA. Cells were selected in 500 μ g/ml G418 for two weeks. Monoclonal cell lines were made by picking individual colonies and identifying those with high surface CD8 expression by staining with phycoerythrin-conjugated anti-CD8 antibody (BD Biosciences, San Jose, CA) (293 cells normally do not express CD8). Polyclonal cell lines were made by purifying a population of cells using Miltenyi anti-CD8 microbeads and a MACS mini-column (Miltenyi Biotec, Auburn, CA). Gene targeting was measured by transfecting 293/A658 cells with RS2100 with or without a Sce expression plasmid

along with a control plasmid (pON405) to determine the transfection efficiency. We used three different promoters to drive Sce expression: PGK, cytomegalovirus ("CMV"), and CBA. The cells were then incubated for 3 days and the percentage of GFP positive cells measured by flow cytometry using a FACScan (BD Biosciences, San Jose, CA). The gene targeting rate was determined by normalizing the measured percentage of GFP positive cells to the transfection efficiency. B) Representative Flow Cytometry Plots of Gene Targeting. GFP positive cells were quantitated in region "R2" as depicted in the right flow plot. The left plot, "S-GT," shows 293/A658 cells after transfection with RS2100 alone. The two GFP positive cells are circled and represent spontaneous gene targeting events. The middle plot, "DSB-GT," shows 293/A658 cells after co-transfection with RS2100 and CBA-Sce. In this plot there are numerous GFP positive cells in region R2. The right plot shows 293/A658 cells after transfection with Sce expression plasmid alone. There are no GFP positive cells. C) Gene Targeting Rates in 293 Cells. The results are shown as both the number of gene targeting events per million transfected cells ("Events/ 10^6 cells") plus/minus one standard deviation and as an overall rate. The results are shown for four different gene targets. In the "1bp mutation" target, a nonsense mutation was created in the GFP gene at bp 321 of the coding region that abrogates functional GFP expression. For the "7 bp insertion" target, a 7 bp sequence was inserted at bp 327 of the GFP coding region. The gene target for the "35 bp insertion" was A658 and the target for the "66 bp insertion" was QQR8 (schematized in Figure 1B). The row labeled "Sce" shows whether Sce was co-transfected or not. The column labeled "Fold Stimulation by Sce Induced DSB" was the stimulation of the

gene targeting rate on target A658 induced by expression of Sce. D) Time Course of Gene Targeting. The relative rate of DSB-GT was normalized to day 3.

Supplementary Figure 2

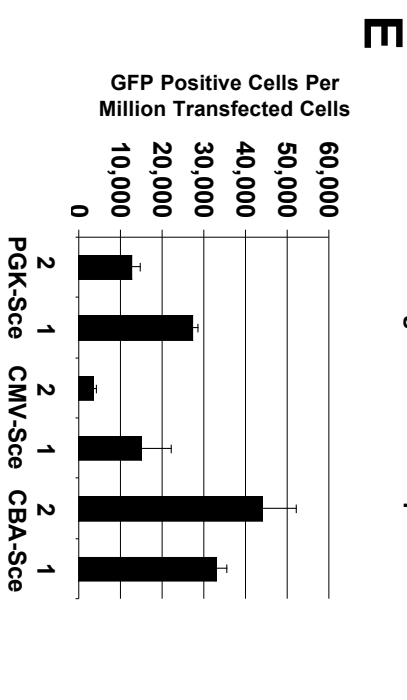
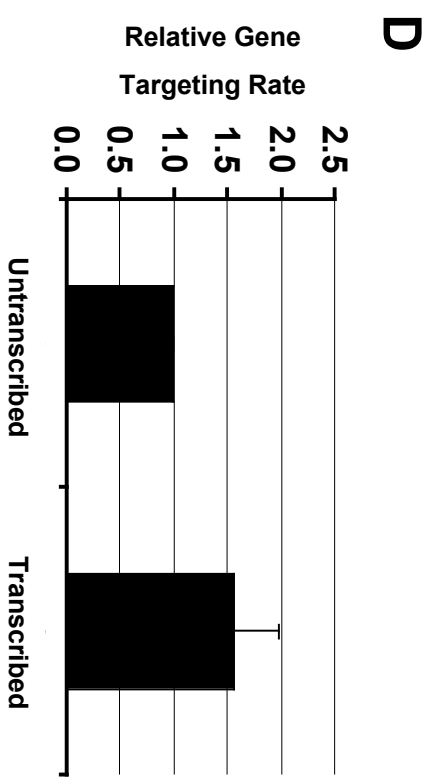
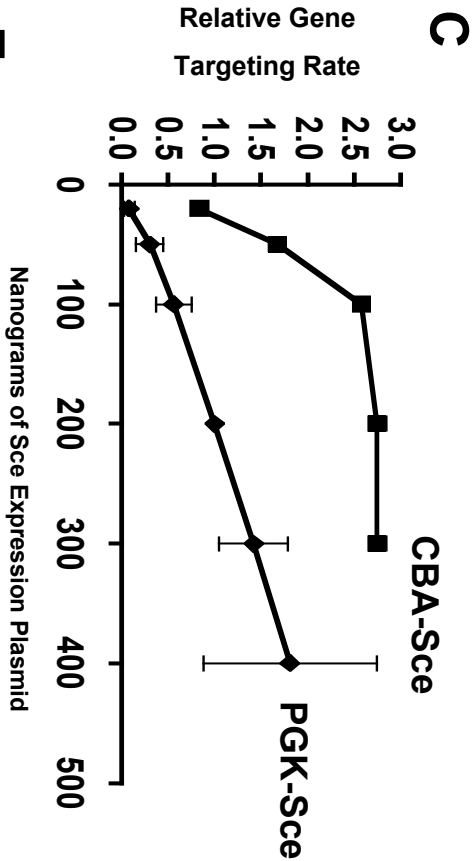
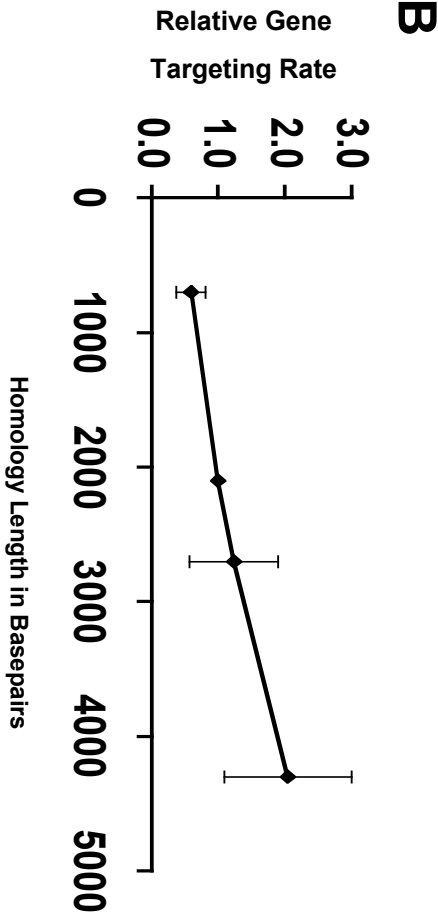
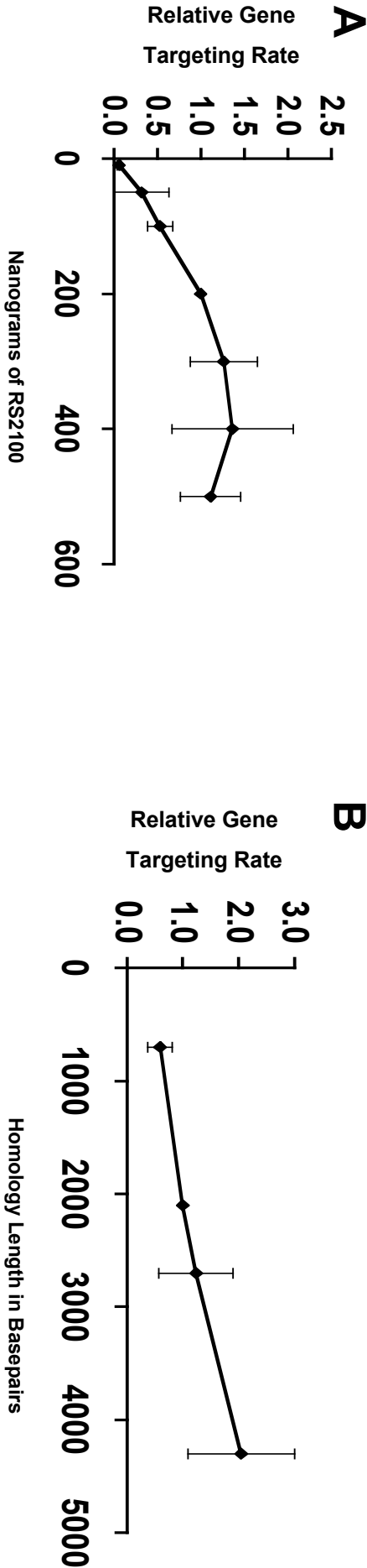


Fig. S2. Parameters regulating the rate of DSB-induced gene targeting. In these experiments, transfections were performed by the calcium phosphate technique in 24-well plates. In plots A-D, the rates of gene targeting were normalized to the standard conditions of using 200 nanograms (ng) of RS2100 and 200 ng of PGK-Sce. In experiments where the amount of a transfected component was varied, the total DNA amount was kept constant by adding pBSK(-) plasmid (Stratagene, La Jolla, CA). A) Gene Targeting Rate vs. Substrate Amount. The results were normalized to the rate of gene targeting obtained after transfecting 200 ng of RS2100. B) Gene Targeting Rate vs. Homology Length. The results were normalized to the rate obtained with plasmid RS2100. The plasmid with 800 bp of homology (RS800) was missing the IRES-CD8 component of RS2100. The plasmid with 2700 bp of homology (RS2700) is depicted in Fig. S1A. The plasmid with 4200 bp of homology (RS4200) had the addition of both the WPRE and the PGK-NEO components to the 3' end of RS2100. A constant amount (200 ng) of each repair substrate was transfected but the relative rate of DSB-GT was normalized to the molar amount transfected. C) Gene Targeting Rate vs. Amount of Sce Expression Plasmid Transfected. The results were normalized to the rate of gene targeting obtained when 200 ng of PGK-Sce was transfected. D) Gene Targeting Rate vs. Transcriptional Status of Repair Substrate. "Untranscribed" was the rate of DSB-GT using RS2100. "Transcribed" was the rate of DSB-GT when the sense strand of RS2100 was transcribed using the CMV promoter (CMV-RS2100). The rates were normalized to the rate of gene targeting obtained using RS2100. E) Optimization of Gene Targeting. Columns labeled "1" are when Sce and RS2100 are on the same plasmid and columns

labeled “2” are when Sce and RS2100 are on separate plasmids. 30,000 GFP positive cells per million transfected cells is equivalent to a gene targeting rate of 3%.

Abbreviations: CBA, cytomegalovirus enhancer/chicken beta-actin promoter; PGK, phosphoglycerate kinase promoter; Sce, *I-SceI* endonuclease, CMV, human cytomegalovirus early enhancer/promoter.

References

- S1. C. Deng, M. R. Capecchi, *Molecular and Cellular Biology* **12**, 3365-3371 (1992).
- S2. R. Zufferey, J. E. Donello, D. Trono, T. J. Hope, *J Virol* **73**, 2886-92. (1999).